Tac-Simur: Tactic-based Simulative Visual Analytics of Table Tennis

Jiachen Wang, Kejian Zhao, Dazhen Deng, Anqi Cao, Xiao Xie, Zheng Zhou, Hui Zhang, and Yingcai Wu

Abstract— Simulative analysis in competitive sports can provide prospective insights, which can help improve the performance of players in future matches. However, adequately simulating the complex competition process and effectively explaining the simulation result to domain experts are typically challenging. This work presents a design study to address these challenges in table tennis. We propose a well-established hybrid second-order Markov chain model to characterize and simulate the competition process in table tennis. Compared with existing methods, our approach is the first to support the effective simulation of tactics, which represent high-level competition strategies in table tennis. Furthermore, we introduce a visual analytics system called Tac-Simur based on the proposed model for simulative visual analytics. Tac-Simur enables users to easily navigate different players and their tactics based on their respective performance in matches to identify the player and the tactics of interest for further analysis. Then, users can utilize the system to interactively explore diverse simulation tasks and visually explain the simulation results. The effectiveness and usefulness of this work are demonstrated by two case studies, in which domain experts utilize Tac-Simur to find interesting and valuable insights. The domain experts also provide positive feedback on the usability of Tac-Simur. Our work can be extended to other similar sports such as tennis and badminton.

Index Terms—Simulative Visual Analytics, Table Tennis, Design Study

1 Introduction

Simulative analysis plays an important role in competitive sports. For example, it has been used in basketball [31] and tennis [35] to provide prospective insights for coaches to improve the performance of players in future matches. In table tennis, simulative analysis also helps obtain valuable insights into the tactical behaviors of players [24, 38, 44]. However, prior studies mostly used complicated mathematical models for simulative analysis without adopting any visual or interactive means. According to our experts, they experience difficulties in exploring model spaces and determining meaningful patterns. Similarly, coaches and players face challenges in understanding the analysis results. Thus, simulative analysis has seldom been accepted or adopted by professional teams in table tennis. In this work, we take a visual approach in contrast to the aforementioned simulative analyses and conduct a design study in the simulative visual analytics of table tennis.

Simulative analysis in table tennis can obtain the outcomes of future matches [24, 38, 44], which is similar to predictive analysis. However, in addition to the outcomes, simulative analysis also focuses on the adequate modeling of each step within the simulated process [8] while predictive analysis only attaches importance to the accuracy of the predicted results [17]. Therefore, we cannot directly apply the existing methods of predictive visual analytics to the simulative visual analytics of table tennis. In sports visualizations, iTTVis [40] is a representative visual analytics system for investigating table tennis data. However, iTTVis is used to investigate the correlation of statistics in a table tennis match without any prospective insights into future matches, which is different from a simulative visual analytics system. Therefore, a newly-designed visual analytics system for the simulative analysis of table tennis is highly demanded by our domain experts.

We worked closely with our domain experts, who are data analysts working for one of the top national table tennis teams around the world, to develop such a visual analytics system. During the collaboration, we encountered two major challenges. The first challenge is to construct an effective model for the tactic-based simulation of table tennis. Existing studies mostly use a first-order Markov chain model for simulation [24, 38, 44], which cannot simulate the effects of tactics. In table tennis, a tactic represents three consecutive strokes (the action of hitting the ball once is defined as a stroke) according to the experts. Therefore, the effect of a tactic reflects upon the correlation among three consecutive strokes. However, the first-order model can only cover the correlation between two adjacent strokes given the Markov property, leading to ineffective simulation of matches. The second challenge is to provide effective exploration and explanation for simulative analysis. Experts need to conduct diverse simulation tasks to explore possible optimization strategies for future matches. However, a large number of strokes and tactics can easily overwhelm experts during exploration. Moreover, according to the experts, they require the simulated match process to interpret the results but they are often confused with the process depicted by complicated mathematical models. The transition from an abstract model to practical competition data is necessary. Therefore, providing friendly interactions for strategy exploration and effective visualizations for result explanation is highly challenging.

We propose a hybrid second-order Markov chain model to solve the first challenge. We extend the classical technique (i.e., three-phase method [39]) for tactical analysis in table tennis and integrate it into a second-order Markov chain model [5]. The new model can involve the effects of tactics in the simulation process and obtain a more accurate result compared with the first-order Markov chain model. As for the second challenge, we design Tac-Simur, a visualization system for tactic-based simulative analysis. The system contains three components: a navigation component, an exploration component, and an explanation component. The navigation component enables experts to browse the matches of multiple players and locate the players and tactics of interest. Experts can then explore the simulation process and evaluate the effects and feasibility of simulations in the exploration component. Lastly, experts can explain the simulation result in the explanation component through the effective visualization of the simulation model.

Contributions of this work are as follows:

- We **characterize the problem domain** of the simulative visual analytics of table tennis into three aspects, namely, navigation, exploration, and explanation.
- We **introduce a well-established model** to adequately simulate the table tennis matches on the basis of tactics.
- We **develop a visual analytics system** to help facilitate the simulative visual analytics of table tennis.

2 Related Work

In this section, we introduce prior studies that are closely related to our work, including tactical analysis of table tennis, prospective analysis in sports, visual analytics in sports, and predictive visual analytics.

- *J. Wang, K. Zhao, D. Deng, A. Cao, X. Xie and Y.Wu are with the State Key Lab of CAD&CG, Zhejiang University. E-mail: {wangjiachen, zhaokejian, dengdazhen, caoanqi, xxie, ycwu}@zju.edu.cn. Y. Wu is the corresponding author.*
- *Z. Zhou and H.Zhang are with the Department of Sport Science, Zhejiang University. E-mail: { zheng.zhou, zhang_hui}@zju.edu.cn*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2.1 Tactical Analysis of Table Tennis

Tactics are one of the focal points in table tennis analysis, and the number of studies in this field has increased in recent years [8]. Tamaki et al. proposed a time-saving method to explore scoring rates in table tennis games based on shot numbers [30]. Zhang et al. also proposed a representative data mining method based on the classical three-phase method [39] to assess the technique effectiveness of table tennis games [45, 47]. Pradas et al. presented a temporal analysis approach to evaluate different temporal structures of table tennis games [27]. Lanzoni et al. analyzed the technical and tactical differences of table tennis players in three categories [13]. These studies primarily used the statistics of key indicators to evaluate player performance in matches. Moreover, techniques such as data mining and artificial neural networks also have been widely applied to tactical analysis for identifying association characteristics in table tennis strokes [46]. However, these methods cannot provide prospective insights for experts. Hence, both Zhang [44] and Pfeiffer et al. [24] proposed a first-order Markov chain model for tactical analysis based on simulation of table tennis games. Wenninger and Lames further improved the model with a numerical derivation method, which aimed to identify the effect of different tactical behaviors on the scoring rate [38]. However, this model cannot simulate the effects of tactics, which motivates us to develop a new model to integrate the tactic into the simulation.

2.2 Prospective Analysis in Sports

Recently, the interest in prospective analysis of competitive sports has increased significantly [1]. For example, Vračar et al. integrated the Markov process with multinomial logistic regression to predict points in basketball matches [31]. In addition, several methods, such as the network-based prediction model [6], the computational random-walk model [9], and many other models [3, 18, 21], predict the future outcomes of different sports. However, these methods primarily focused on the outcome of matches and disregard the complexity of the inner process of matches. Wei et al. developed a series of methods [33–35] to simulate tennis matches and reveal patterns within the match process. However, these models cannot be applied to table tennis due to distinct data structures and match rules. Apart from tennis, the Markov chain model that simulates table tennis matches is also a significant method proposed by Zhang [44], Pfeiffer et al. [24], and Wenninger and Lames [38]. This model inspires our work through its simulation of the state-transition feature in table tennis matches. However, this approach fails to effectively model tactics due to the Markov property. Therefore, we propose a new simulation model based on this model.

2.3 Visual Analytics in Sports

Visual analytics in sports has elicited considerable research attention and the number of related works is increasing rapidly [23]. In basketball, Cervone et al. proposed POINTWISE to predict points using player-tracking data in NBA games [2]. Meanwhile, other studies of basketball such as Courtvision [10], GameFlow [4], MatchOrchestra [32], and BKVis [16] focused on shot distributions, events, and performance indicators, respectively. In soccer, SoccerStories provided an innovative visualization interface for exploring soccer games with flexible interactions [22]. ForVizor used a tailored Sankey diagram to represent changes in team formations in soccer games [41]. In baseball, StatCast Dashboard [11], Baseball4D [7], and Baseball Timeline [20] were representative works for the visual analytics of baseball tracking data. Furthermore, there were also a large number of significant studies about other sports events, like rugby [14], ice hockey [25], fencing [48] and tennis [26]. In particular, iTTVis supported the visual analytics of table tennis data [40]. It provided visualizations from diverse aspects such as box scores and strokes attributes. However, these approaches cannot support simulative analysis, thereby motivating us to develop a visualization system to help users explore and explain the simulation.

2.4 Predictive Visual Analytics

We also refer to the works of predictive visual analytics given the similarity between prediction and simulation. Lu et. al conducted a survey regarding the state-of-the-art in predictive visual analytics [17]. However, we cannot directly apply the existing works to simulative visual analytics because simulative analysis focuses on the adequate process simulation and the precise result prediction whereas predictive analysis only focuses on the latter one. A complete review of predictive visual analytics is beyond the scope of this work. More details about predictive visual analytics can be found in the recent survey [17].

3 Background and System Overview

In this section, we firstly introduce definitions of terminologies used in the domain and the data used in this work. Then we analyze the requirements for simulative analysis of table tennis.

3.1 Background

Table tennis is a highly confrontational sport wherein two opposing sides hit a ball back and forth on the table with a net until one side misses the ball and the other scores a point. The structure of a match is illustrated in Fig. 2 black. A formal match is typically a best-of-seven series and involves four to seven games. Each game contains tens of rallies. The winner of each game is generally the one who first wins eleven rallies. Each rally contains strokes hit by two players.

A stroke is an action wherein a player hits the ball once with his/her racket. It is the basic observation unit during data collection [24]. We use three technical attributes shown in table 1 to describe a stroke, namely, stroke placement, stroke technique, and stroke position.

A rally is the composition of all strokes given by two players to score one point. It is the basic unit for judging one point. Therefore, it is the basic unit of analysis during match simulation, and the simulation of a rally can be regarded as the simulation of a match [24].

A tactic is composed of any three consecutive strokes in a rally. As Zhang and Zhou [47] indicate, the attribute of the current stroke largely depends on the former two strokes from a player himself/herself and his/her opponents. Specifically, the first stroke of a player is the inducement of a tactic. If a tactic is successful, then his/her opponent would hit back the first stroke with the second stroke expected of him/her, and then the player would hit back with the third stroke he/she prepared in advance with a high scoring rate. That is, if a tactic works, the player are more like to seize the initiative and win the rally.

A tactical adjustment is a strategy to change the usage of tactics for a player. Experts typically conduct a tactical adjustment by changing the technical attributes (Table. 1) of strokes. For example, they will increase the proportion of using reverse (a stroke technique) in serve rounds to determine whether such an adjustment can increase the winning rate. Experts will also concurrently change multiple stroke attributes during simulation.

3.2 Data Description

The data is manually collected from match videos by professional table tennis players. Both the technical attributes of strokes and the contextual data such as the maker of a stroke, the order of all strokes, and the score information are included during collection. The primary stroke attributes used for analysis are presented in Table. 1 as follows.

Table 1. The stroke attributes

Stroke placement	Position of the ball on the table tennis table after it is hit (*i.e., short forehand, short middle, short backhand, half-long forehand, half-long middle, half-long backhand, long forehand, long middle, and long backhand*).
Stroke technique	Technique used to hit the ball (*i.e., pendulum, reverse, tomahawk, topspin, quick attack, smash, flick, twist, push, short, slide, block, and lob*).
Stroke position	Position of the player when he/she is hitting the ball (*i.e., forehand, backhand, backhand turn, and pivot*).
Stroke player	Player hitting the ball.
Score A/B	Winner of the rally a stroke belongs to.
Match ID / Stroke ID	Index of the match / stroke.

3.3 Requirement Analysis

We worked with two domain experts, namely, a professor and his Ph.D. candidate from the Department of Physical Education. Both were former professional table tennis players and have worked for one of the top national table tennis teams in the world for more than 5 years. Besides, one of our co-authors who majors in computational sports science worked as a Liaison [29] during the cooperation. At first, we identified the problem domain with a simple pilot system. Then, we developed the simulation model through trial and error based on the problem domain. Finally, we iteratively designed and developed the visual analytics system according to the experts' feedback and suggestions. The detailed milestones are as follows.

◇ **Characterizing the problem domain.** We held weekly meetings with the domain experts to identify the problem domain of the simulative analysis of table tennis. To facilitate this process, we developed a pilot system to help identify the limitations of existing methods and collect the requirements. After two-month exploration and iteration, we finalized the problem domain.

◇ **Developing the simulation model.** We tried various simulation models and it turned out that the simulation process of the Markov chain model is more comprehensible than deep learning models. Therefore, we constructed our model based on the second-order Markov chain model and the typical theories of table tennis.

◇ **Designing interactive visualizations.** When the model was ready, we started the visual design of the system according to the requirements and the model. We designed the initial system at first and iterated the design with experts during weekly meetings.

◇ **Developing the analytic system.** We developed a prototype based on the design and deployed it on the web for the experts to use. During this period, the interactions and visualizations were frequently revised in accordance with the experts' feedback. The system was iteratively refined through considerable revisions.

We summarized the requirements from the experts into three aspects, namely, navigation, exploration, and explanation.

Navigation can help experts have an overall idea about the players they are going to analyze, facilitating the analysis process.

R1 *What are the matches and the results of each player?* When experts select the data to be analyzed, they will firstly choose a player they are interested in and review the result of the matches he/she participated in. Usually, they tend to analyze the close matches or the ones revealing the characteristics of the player.

R2 *What is the playing style of a player? What is the key tactic in the matches?* The playing style of a player is intrinsically indicated by the type of tactics he/she used most, and the type of tactics he/she scored most. This information can provide navigation and reference for analysts while applying adjustments. For example, if a player is poor at a frequently-used tactic A, then the experts will expect to improve this tactic by adjusting its strokes. Moreover, the information is also necessary to explain the simulation results.

Exploration can help experts integrate empirical domain knowledge into the tactical adjustments and enhance the adjustment feasibility.

R3 *What kind of strokes is worth adjusting most?* A player could have lots of strokes that can be adjusted. However, instead of adjusting each stroke to examine the result, experts would like to find a set of key strokes that may be important for improving the performance. This can help them significantly reduce the time for searching appropriate adjustment strategies.

R4 *What is the effect of an adjustment strategy? What is the feasibility of an adjustment strategy?* To evaluate an adjustment strategy, experts need to know the effect and feasibility of the adjustment strategy. Specifically, the effect means the increment/decrement imposed to final winning rates and the feasibility characterizes the difficulty of utilizing an adjustment strategy in real scenarios.

Explanation can help experts understand the simulated match process, interpret the results of adjustments, and communicate the findings to players to improve the performance.

R5 *How does a tactical adjustment influence the strokes and tactics?* Experts need to know the reasons for the positive or negative effect of a particular adjustment. Generally, the reasons lie in the influence of the adjustment to the latter strokes. For example, the positive effect of increasing the usage of quick attack, an offensive stroke technique at the second stroke is because this adjustment can further raise the usage of other offensive techniques at latter strokes, which can easily enhance the winning rate of the player.

R6 *How to conduct a tactical adjustment in practice?* Once an adjustment strategy is discovered, experts will expect to figure out how to conduct it in real scenarios so that they can communicate it to the players. Specifically, experts need to examine the relationship between the adjusted stroke and the former strokes to provide a practical solution for players (e.g., to increase the usage of quick attack at the third stroke, you need to use pendulum to serve).

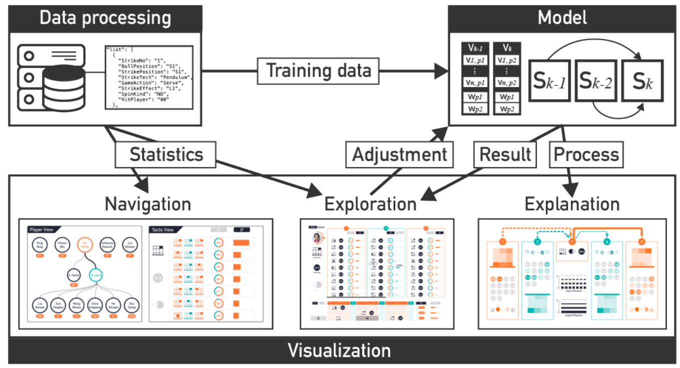

Fig. 1. The overview of the system. The system consists of three components, i.e., the data processing component, the model component, and the visualization component.

3.4 System Overview

Tac-Simur is a web application with three components: the data processing component, the model component, and the visualization component (Fig. 1). The data processing component is responsible for extracting the required attribute data from the raw database and providing data interfaces for modeling and visualization. The model component initializes the hybrid Markov chain model for each player with the data from the data processing component. The visualization component is the interface for simulative visual analytics. This component consists of three parts, namely, navigation, exploration, and explanation. The navigation part helps users locate a dataset and tactics of interest for further analysis. The exploration part supports flexible adjustments to the matches. Once an adjustment is conducted, the model component will receive it and send the simulation results back to the exploration part for visual comparison and evaluation. The explanation part provides a straightforward presentation of the simulation process for reasoning and validating the results generated in the exploration part. The data processing component is implemented via MongoDB and Express.js. The model component is implemented via Python. The visualization component is implemented via Vue.js.

4 Model for Simulation

In this section, we firstly define the simulation task in table tennis and briefly introduce the Markov chain model. Thereafter, a well-established hybrid second-order Markov chain model is presented to address the limitations.

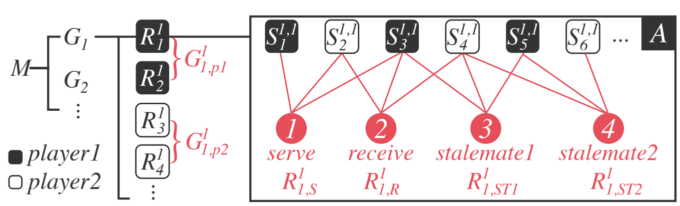

Fig. 2. The example of the structure of a match. M refers to "Match", G_i refers to "Game", R^i_j refers to "Rally", and $S^{i,j}_k$ refers to "Stroke". The red parts are newly added compared with the old model. (A) illustrates the division of the four phases.

4.1 Task Definition and Model Overview

As shown in (Fig. 2 black), a table tennis match is a set of games $M = \{G_1, G_2, \dots, G_{n_M}\}$ (for simplicity, we use the notation n_* to indicate the amount of the elements in $*$) and a game consists of a number of rallies $G_i = \{R^i_1, R^i_2, \dots, R^i_{n_{G_i}}\}$. For each rally, it can further be quantified as a stroke sequence $R^i_j = \{S^{i,j}_1, S^{i,j}_2, \dots, S^{i,j}_{n_{R^i_j}}, P^i_j\}$, where $S^{i,j}_k$ is the k^{th} stroke and P^i_j is the result of R^i_j (Player 1 wins or loses). $S^{i,j}_k, P^i_j \in \mathbb{S}$, a discrete state space characterized by the combination of attributes in Table 1. Therefore, the simulation of a table tennis match intrinsically involves simulating the stroke sequences (rallies). Specifically, given the first stroke, $S^{i,j}_1$, a simulation model should obtain the remaining strokes including the winner P^i_j by assigning the states from $\mathbb{S}$ to them.

The original model [24] for simulation of matches stems from Martin Lames' Markov chain model [12]. However, when applied to the simulation task, it has two limitations:

Tactic modeling is inadequate. The first-order Markov chain model treats each rally within a game as a first-order Markov process, in which the S_k is only determined by S_{k-1} and a static transition matrix T(Fig. 3(C) blue). It cannot simulate the table tennis matches precisely because a tactic is defined by the states of three consecutive strokes (i.e., $T_k = (S_k, S_{k+1}, S_{k+2})$ and $S_k, S_{k+1}, S_{k+2} \in \mathbb{S}$). The transition probability between two states from $\mathbb{S}$ varies in different phases of a rally (e.g., serve, receive, and two stalemates Fig 2(A)). Thus modeling the tactic with a one-step transition matrix is inadequate.

Stroke characterization is insufficient. In table tennis, a stroke is typically characterized by *stroke placement* (*Pla*), *stroke technique* (*Tec*), and *stroke position* (*Pos*) (Table. 1). The original model characterizes the transition of stroke states with only one of the three attributes (Fig. 3(B) blue), which is insufficient to simulate the real situation.

However, compared with other deep learning models, this model is more comprehensible because it depicts the interaction between players clearly through a transition matrix. As a simple probabilistic graphical model, the Markov chain can trace and interpret each step within the simulation of table tennis as a practical stroke within the match.

Therefore, based on the original model, we constructed a **hybrid second-order Markov chain model**. The new model improves the two aspects mentioned above. First, we split all rallies and divided them into four kinds of Markov processes on the basis of the three-phase method [39] (Fig. 2(A)). Each Markov process consists of two transition matrices (Fig. 3(A) red). One depicts the transition from S_{k-1} to S_k and the other depicts the transition from S_{k-2} to S_k. In this way, tactics can be modeled properly as the current stroke is determined not only by the former stroke but also by the one before the former stroke (Fig. 3(C) red). Thus the simulation process of the hybrid second-order Markov chain model is more precise and reasonable. Second, we expanded the number of attributes used in stroke characterization to a maximum of three, namely, $S_k = (s^k_1, s^k_2, s^k_3)(s^k_1 \in Pla, s^k_2 \in Tec, s^k_3 \in Pos)$ (Fig. 3(B) red). Experts can conduct more complicated and flexible adjustments with strokes characterized by more attributes.

4.2 Hybrid Second Order Markov Chain Model

Here we introduce our hybrid second-order Markov chain model. The domain experts indicated that players always exhibit tactical awareness during matches. Therefore, a stroke is highly relevant to the former two strokes as a tactic contains three consecutive strokes. To model this feature, we ultimately decided to use a higher order Markov chain model [5]. Briefly, an n^{th}-order Markov chain model is composed of n first-order models. Therefore, a conventional n^{th}-order Markov chain contains $(m-1) \cdot m^n$ parameters (m is the number of state attributes). Raftery [28] provided an optimization method to simplify the computation process. Ching et al. further [5] extended this method to a more accurate one. Here we adopted the concept of the second-order Markov chain model to construct a hybrid model as follows:

$$V_k = \lambda_1 \cdot V_{k-1} \cdot T_1 + \lambda_2 \cdot V_{k-2} \cdot T_2, \tag{1}$$

where $\lambda_1 + \lambda_2 = 1$. Given that a tactic involves three strokes, a second-order Markov chain model (Equation. 1) can appropriately model the influence of tactics. Here T_1 denotes the transition matrix from the $(k-1)^{th}$ stroke to the k^{th} stroke whereas T_2 indicates that from the $(k-2)^{th}$ stroke to the k^{th} stroke (Fig. 3(A) red). λ_1 and λ_2 are the weights estimated by minimizing the prediction deviation (Fig. 3(C) red). The transition matrices, T_1 and T_2, of the model is an empirical transition probability matrix. As shown in Fig. 3(A), the row headers and column headers of the matrices T_1 and T_2 (T for first-order Markov chain model) are the states of a former stroke and a latter stroke respectively. The transition matrices and initial state vectors are estimated by the frequency of each state from the analyzed matches. Moreover, the different phases in a rally are simulated by different Markov processes. In our case, the model comprises four Markov processes, as shown in Fig. 2(A). According to Wu et al. [39], the playing strategy of a player varies over rallies. When the player serves, he/she tends to conduct more offense than defense. However, when he/she receives, the situation is reversed. Meanwhile, the first tactics of the two players in a rally are more diverse than those afterward. Therefore, we first subdivided the rallies of a game ($G_i = \{R^i_1, R^i_2, \dots, R^i_{n_{G_i}}\}$) into two subsets ($G_{i,P1}$ and $G_{i,P2}$) according to the serving player (Fig. 2 red). Further, we split a rally (R^i_j) into four phases, the serve phase ($R^i_{j,S}$), the receive phase ($R^i_{j,R}$), and the stalemate phases of two players ($R^i_{j,ST1}$ and $R^i_{j,ST2}$)(Fig. 2(A) red). $R^i_{j,S}$ and $R^i_{j,R}$ are the first tactics of both players, namely, the serve tactic for one player and the receive tactic for the other, and $R^i_{j,ST1}$ and $R^i_{j,ST2}$ include the tactics at stalemate phases.

During simulation, two initial state vectors, namely, V_1 corresponding to the first stroke S_1 and V_2 corresponding to the second stroke S_2 are required. Then the vectors of strokes (V_3 and V_4) can be obtained by Equation. 1 with the transition matrices in R_S and R_R. After that, with V_3 and V_4, the vectors of the strokes in R_{ST1} and R_{ST2} are further obtained, as well as the convergent results.

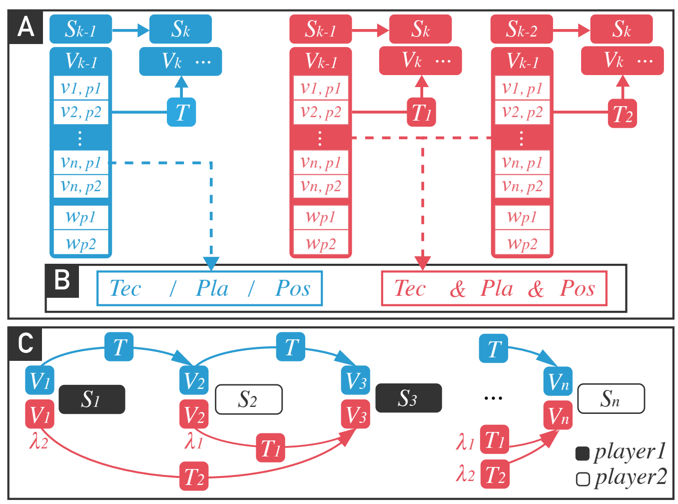

Fig. 3. The original model and the new model. Blue encodes the original model, red encodes the new model. (A) and (B) present the state vectors. Specifically, in (A), S_k represents the stroke, V_k represents the state vector of each stroke S_k, $v_{i,p1}$ and $v_{i,p2}$ represents the state value of stroke attributes and w_{p1} and w_{p2} represents the state value of winning of the two players. (C) presents the simulation process.

4.3 Making Adjustments

In the original model, adjustments are achieved by tuning the empirical transition matrix T. Then changes in the final scoring rates quantify the effects for evaluation. However, according to the experts, tuning the transition matrix means changing the playing style of a player while giving each of his/her stroke in a rally. It is too difficult for players to achieve the adjustments. Therefore, we tune the state vectors in the new model. The tuned vector is calculated based on all of the probability values in the vector. Specifically, if we want to adjust Player 1 (P1) and need to adjust the state value $v_{i,P1} \in V_k$, where $v_{i,P1}$ is the probability value of strokes given by flick to long-forehand with forehand, then we need to enhance the utilization rates of tactics that contain the aforementioned stroke at S_k. We intially increase the $v_{i,P1}$ and then decrease $v_{j,P1} (j \neq i)$ in V_k. According to Pfeiffer et al. [24], the function for deflection is as follows,

$$\delta v_{i,P1} = C + B \cdot A \cdot v_{i,P1} \cdot (1 - v_{i,P1}) \tag{2}$$

where $\delta v_{i,P1}$ is the change in probability; C is a constant that describes the deflection in the border probability; B is a constant that describes the maximum value of the relative magnitude of deflection; and A is a normalization factor that allows the constant B to be equal to the maximum deflection value. In the current work, the constants are set to $C = 0.05$, $B = 0.25$ and $A = 4$ based on the previous work [24] to keep the deflection value between $1\% \sim 6\%$. The compensation function is:

$$\delta v_{j,P1} = -(v_{j,P1}/(1 - v_{i,P1})) \cdot \delta v_{i,P1} \tag{3}$$

where $\delta v_{j,P1}$ is the change in the probability.

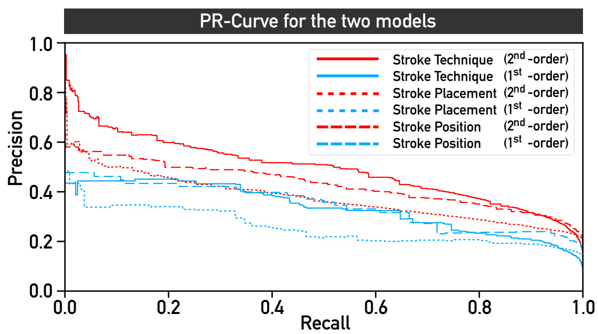

Fig. 4. The evaluation between the two models. The red lines represent the new model and the blue ones represent the old model.

4.4 Model Evaluation

A Markov chain model computes the distribution of the states at every step, thus given the sequential data, the first-order Markov chain and our hybrid second-order Markov chain are supposed to predict the following states accurately. We regard it as a multi-classification problem. At every step, we take the state with the highest probability as the classification result and compare it with the benchmark. The results of the two models with a single attribute are shown in Fig. 4. The Precision-Recall Curve is adopted to evaluate the performance of the model among all states (classes). A larger area enclosed by the PR Curve, the x-axis, and y-axis reveals a better performance. From the figure, we find that with the same precision, our model obtains higher recall rates; with the same recall rate, our model has higher precision. For *stroke technique* (*Tec*), *stroke placement* (*Pla*), and *stroke position* (*Pos*) (Table 1), respectively, our model outperforms the original first-order Markov chain model.

5 Visual Design

In this section, we first introduce the design goals that guide our system design. Then, on the basis of these goals, we illustrate the visual encoding and interaction in our system.

5.1 Design Goal

We further derive the following design goals for the visual design in accordance with the requirements summarized before

G1: *Visual organization of all matches for overview* (**R1**). Since the result and players of a match are the most important information for experts to select matches to be analyzed(R1), an overview of all matches specified by the match score and characteristics of the players involved can facilitate the identification of matches of interest.

G2: *Visual sorting and filtering of tactics for navigation* (**R2**). The experts tend to pay attention to the scoring rates and utilization rates of the tactics because these kinds of information can help find the tactics of interest and depict the playing style of a player during navigation (R2). Given the strength of sorting in decision making [15, 36, 37], we employ techniques of visual sorting, as well as filtering to facilitate navigation of these scoring rates and utilization rates .

G3: *Visual enumeration of stroke for exploration* (**R3**). The exploration space of adjustments is enormous due to the diversity of strokes. Experts hope to explore the adjustments flexibly and efficiently. Therefore, the system should visually enumerate all optional adjustments in a particular order for experts to choose freely and support convenient sorting and filtering techniques to facilitate exploration.

G4: *Visual recording of adjustments for evaluation* (**R4**). The effect of an adjustment strategy is important for evaluation (R4). Experts hope to simultaneously evaluate both the effect and feasibility of each strategy. Besides, they also require to compare different strategies to identify the optimal or practical one. Therefore, it is necessary to have an independent view to save the applied adjustments for subsequent evaluation and comparison.

G5: *Visual illustration of correlation for explanation* (**R5, R6**). The correlation among strokes within the match process can reveal the interactions between two players. This information is indispensable for experts to comprehend the effects of adjustments (R5) and further explain the adjustments to players (R6). However, given the multivariate stroke attributes and complicated composition of tactics, experts always exert considerable effort to examine the correlation. Therefore, the effective and efficient visual illustration of the correlation is necessary.

G6: *Representative icons of strokes and tactics* (**R2, R3, R4, R5, R6**). Icons are effective for depicting physical objects and concepts [19]. Therefore, icon-based visualizations can facilitate exploration of strokes and tactics along with their utilization rates and scoring rates (R2, R3, and R5). In this manner, the adjustments can be evaluated and compared efficiently (R4), and the coaches and the players can easily understand the processes and results(R6).

5.2 System Design

In Tac-Simur, we design a player view for match navigation (R1), a tactic view for tactic navigation (R2), and a simulation view for exploration of optional adjustments (R3), evaluation of adjustment strategies (R4), and explanation of the results of adjustments (R5, R6).

In the player view, experts can select the matches of interest through the overview of all matches for analysis (G1) (Fig. 5(A)). Then, experts can navigate the tactics and locate those of interest for subsequent analysis by sorting and filtering in the tactic view (G2) (Fig. 5(B)). Based on the selected tactics, experts can apply diverse adjustments to strokes in the exploration component of the simulation view (Fig. 5(D)) (G3) and record the adjustments they have tried in the evaluation component for evaluation and comparison (Fig. 5(E)) (G4). To figure out the influence and implementation of a particular adjustment, experts can obtain the simulated competition process from the explanation component of the simulation view (Fig. 6(A)) (G5). To facilitate analysis, we employ metaphor-based icons to encode stroke attributes (G6).

We choose orange and cyan to represent the player to be analyzed and his/her opponents as unified color encoding, respectively. The detailed design for each view is described as follows.

5.3 Player View

According to the experts, when they analyze table tennis matches, they usually first target a player and focus on the matches of him/her. Furthermore, the experts tend to group together the matches where the player's opponents have the same handedness to draw highly specific conclusions. Therefore, organizing the matches in a two-level hierarchical structure (i.e., the first level is player and the second level is handedness) can appropriately support such progressive navigation. Given the simple topology of the hierarchical structure, we employ a node-link tree rather than space-filling methods such as a treemap to present the hierarchical structure of matches in the player view (G1) (Fig. 5(A)). Each node in the tree is encoded by a circle with the players' names or category names with a winning rate under it. Initially, the player view horizontally shows all the players in the database for experts to choose (Fig. 5(A1)). After experts choose a player, the corresponding circle will turn to orange and become the root of a tree. The two children of the root are left-handed players and right-handed players (Fig. 5(A2)), respectively. Experts can continue selecting left-handed players or right-handed players to show details regarding the

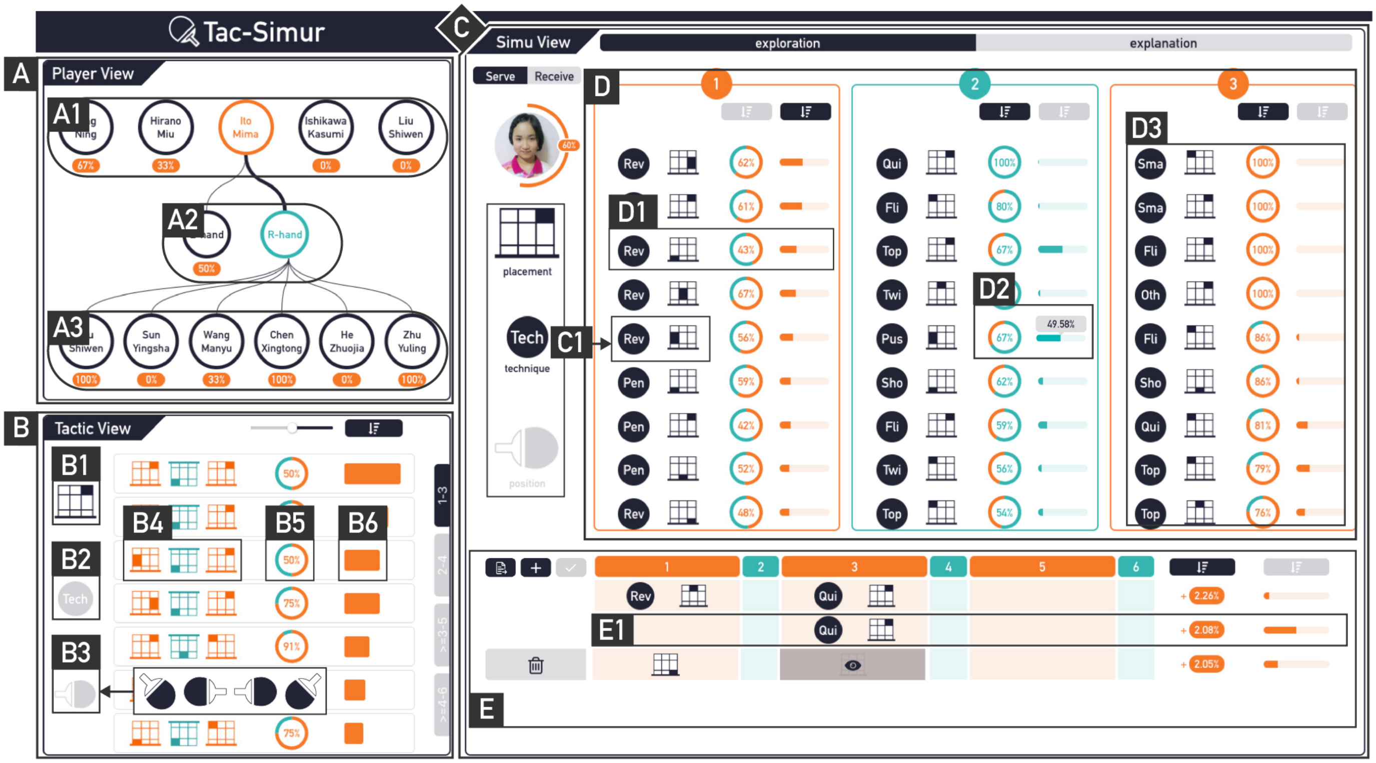

Fig. 5. The system interface. The donut charts and the pie charts encode the scoring rates and utilization rates, respectively. (A) is the player view which displays all matches of Ito Mima and her opponents. It provides navigation of matches. (B) is the tactic view which displays the tactics specified by stroke placement at the serve phase. It provides navigation of tactics. (C) is the simulation view which is under the exploration mode. It contains (D), the exploration component for implementation of adjustments and (E), the evaluation component for evaluation of adjustments. (D) displays all of the adjustment options specified by stroke placement and stroke technique. (E) displays the three optimum strategies generated by the system.

opponents as the leaves of the treemap (Fig. 5(A3)). After the experts choose the player and his/her opponents of interest, the corresponding matches will be selected for analysis.

5.4 Tactic View

The tactic view is a tactic list consisting of all the tactics used by the targeted player in the selected matches. These tactics are presented by icons of strokes (G6). They can be sorted based on scoring rates or utilization rates (G2) and classified according to rally phases.

Icon. iTTVis [40] has provided a set of icons for the three stroke attributes. Although the icons are well-designed, it is difficult for experts to distinguish between the icons of the same attribute (e.g., stroke position) when the visual space is limited (the analysis requires experts to inspect numbers of strokes). Therefore, we revised these icons based on the features mentioned in [19]. The icons of each stroke attribute are as follows.

- ***Stroke placement.*** Table tennis table is widely used to illustrate the ball position of a stroke. Therefore, we use a half table tennis table to encode the stroke placement, which meets familiarity, concreteness, and meaningfulness [19]. Unlike that of iTTVis, the half table is divided into nine grids based on the real scale of the division criterion provided by the domain experts (Fig. 5(B1)). The filled grid represents a certain stroke placement of a stroke.
- ***Stroke technique.*** Encoding the fourteen stroke techniques by color or other visual channels is difficult. Therefore, we follow iTTVis and use the abbreviation of the technique name to represent the stroke techniques (Fig. 5(B2)). This encoding meets meaningfulness and semantic distance [19].
- ***Stroke position.*** In iTTVis, this attribute is encoded by the specific poses of players. However, this encoding is difficult to recognize given the small size of the icon. Therefore, we use a racket to encode the stroke position since different poses can be represented by different means to swing the racket. This encoding meets concreteness and meaningfulness [19]. The four different directions denote backhand-turn, backhand, forehand, and pivot from left to right, respectively (Fig. 5(B3)).

Tactic list. The list is used for navigating tactics for subsequent analysis. The tactics are all described by one kind of attributes because filtering with more attributes will severely diminish the adjustable candidates in the simulation view, thereby limiting exploration. Each row of the list contains three components: the tactic (Fig. 5(B4)), the scoring rate (Fig. 5(B5)), and the utilization rate (Fig. 5(B6)). The scoring rate of a tactic is symbolized by a donut chart. The utilization rate is presented by a bar chart. These two encodings are straightforward for comprehension and efficient for sorting (G2).

Interaction. Sorting buttons and sliders are set above the two rates to enable descending sorting and filtering. Besides, experts can choose to browse tactics in different phases through the button on the right side. Additionally, experts can click the icons on the left side to select the tactic attribute used to depict the tactics (Fig. 5(B1, B2, B3)).

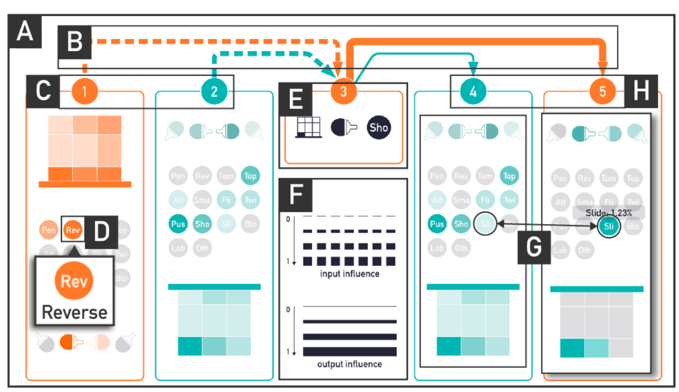

Fig. 6. The explanation mode of simulation view. (B) presents the correlation among strokes. (C) and (H) present the former and latter two strokes of the adjustment (E).

5.5 Simulation View

By selecting tactics of interest in the tactic view, users can conduct adjustments to the selected tactics in the simulation view. The simulation view consists of an exploration component (G3) (Fig. 5(D)), an evaluation component (G4) (Fig. 5(E)), and an explanation component (G5) (Fig. 6). Users can explore potential adjustments with the exploration component and conduct an adjustment with the evaluation component. Explanation of the results of the adjustments is provided by the explanation component. Detailed descriptions are as follows.

The exploration component provides comprehensive options of adjustments based on sorting results of each stroke (G2). This view presents adjustment options for three consecutive strokes (a tactic) at a time because the experts expect to take tactics into consideration while adjusting strokes (Fig. 5(D)). We provide an optional stroke list for each stroke (Fig. 5(D3)). Each item in the optional stroke list consists of three components, namely, the stroke attributes, the scoring rate, and the adjustability (Fig. 5(D1)) from left to right (R3). The icons (G6) and the donut charts are the same as those in the tactic view. The utilization rate is replaced by the adjustability to help assess the feasibility so as to expedite the decision-making process. This value is first calculated by Equation. 2 and then normalized to one. Similarly, all kinds of strokes can be sorted by scoring rates or adjustability coefficients for convenience (G2). The targeted player's avatar is also placed at the left top corner of this view with a winning rate. Once an adjustment option is selected, the winning rate will change accordingly.

The evaluation component records the tactical adjustments made by experts (G4). Each row contains an adjustment strategy and its effect and adjustability (Fig. 5(E1)). Given the rally division method used in our model, the system only supports adjustments before the sixth stroke. The adjustments are recorded by the same icons of the selected in the optional stroke list (G6). The effect is computed by the improved high-order Markov chain model defined in Section 4 and simply shown with the exact value for clarity. The adjustability is calculated by Equation. 2 and demonstrated with a bar chart.

The explanation component explains the manner in which an adjustment can be achieved and the reason why it performs well/poorly (Fig. 6(A)). We simplify the simulation process of the model and present it directly for comprehension purposes. The adjustment is placed at the center of the view (Fig. 6(E)). Only directly related strokes, namely, two former stroke sets(Fig. 6(C)), if they exist, and two latter (Fig. 6(H)) stroke sets are displayed here to avoid information overload. According to the principle of the model, we connect the stroke sets with lines to illustrate the correlation between them (G5) (Fig. 6(B)). The dashed lines represent the influence from the former stroke sets to the current adjustment whereas the solid lines denote the influence from the current adjustment to the latter stroke sets (Fig. 6(F)). The line thickness encodes the weight, λ, of the corresponding influence.

The probable strokes in each stroke set that can affect the adjustment and that are already affected by the adjustment are illustrated. The number of displayed attributes remains consistent with that of the adjustment (Fig. 6(E)). Originally, we plan to enumerate all of the probable strokes of each stroke set. In this way, the most detailed information can be presented to experts. However, this method cannot enable summarization of high-level patterns. According to experts, high-level summaries can be accepted by players more easily. Therefore, we eventually decide to display the correlation among stroke attributes instead of enumerating them with progressive interactions.

The opacity of the icon encodes the influence of the corresponding kind of strokes. For example, Fig. 6(D) illustrates that strokes hit with reverse by the player himself/herself will lead to the adjustment most significantly. Initially, each kind of stroke attribute works separately. If experts click a specific attribute value, then the selected value will be highlighted alone, whereas other kinds of attributes will be filtered (Fig. 6(G)). In this manner, the correlation among stroke attributes can be interactively examined (G5).

Interaction. The interaction in the simulation view is as follows.

- *Changing attributes.* Experts can click the legend of the icon as a selection panel for experts to choose the attributes appearing in the optional stroke lists (Fig. 5(C1)).
- *Sorting options.* Experts can click the two buttons on the top of the optional stroke list to sort strokes.
- *Organizing adjustments.* Experts can generate optimum strategies and add customized strategies through the buttons on the left top of the evaluation component Fig. 5(E). Besides, they can also edit the strategies and investigate each of them.
- *Unfolding details.* Experts can hover on the visual elements in the simulation view for details (Fig. 5(D2)).

6 System Evaluation

In this section, we evaluate the usability of Tac-Simur with two case studies conducted by the experts. We also summarize experts' feedback on Tac-Simur during the interview after the case studies.

6.1 Case Study

The two case studies are based on the data of 12 matches of Ito Mima and 7 matches of Ding Ning. All of these matches are from the high-level events including the *World Cup, World Championship, and ITTF World Tour in 2018*. We deployed Tac-Simur on the web and invited two experts to conduct case studies. Both of the experts are former professional table tennis players. Besides, Experts A is a Ph.D. candidate majoring in the performance analysis of table tennis and Expert B is a professor of sports science. Both experts have collaborated with one of the top national table tennis teams for more than five years. Before the case studies, we introduced the system to the experts. After they got familiar with the encoding and the interaction, they analyzed matches by themselves. We assigned an experimenter to each expert in case the experts forget the encodings and the interactions. The experimenter was only responsible for answering questions about the system.

6.2 Insight 1: Consecutive Quick Attack to Long Backhand is the Key to Serve Rallies for Ito Mima.

This case study tends to improve Ito Mima's performance by adjusting her serve rallies. Ito Mima is one of the top table tennis players whose world ranking was seventh. According to the experts, she is a young player with great potential, thus, they were interested in her matches.

The experts first clicked the circle of Ito in player view (Fig. 5(A)). Then, they respectively checked the left-handed players and right-handed players and found that Ito competed with more right-handed players than left-handed players. Therefore, to improve Ito's overall performance more significantly, they chose all of the Ito's matches with right-handed players for further simulative analysis.

Thereafter, the experts turned to the tactic view to browse Ito's tactics. As they indicated, the tactics used in the serve phase (i.e., the first three strokes of a rally) are the focal points in performance analysis. Therefore, they directly examined the tactics in the serve phase specified by the stroke technique. They sorted all tactics by their scoring rates and only showed those whose utilization rates are more than 1% through the slider to immediately identify the highly significant tactics (Fig. 7(A1)). According to the four tactics (Fig. 7(A2)), the experts found that in most of the tactics, Ito's opponents' technique for receive is push, a control technique whereas Ito's third technique is topspin, an offensive technique. This matched the experts' hypothesis about Ito's playing style. The expert further explained that this means after Ito serves, she tends to attack at first. However, the scoring rates of these tactics were inversely proportional to the utilization rates, which means Ito's performance in the serve phase is poor and can be a breakthrough point for improving her performance. Therefore, the experts further selected these tactics for adjustments.

In the simulation view (exploration mode), the experts expected to adjust strokes from the perspective of stroke placement. They activated the stroke placement and let the system generate the optimum adjustment strategies. The results consisting of a single adjustment indicates that the effect (winning rate) of the adjustment applied to the third stroke (Fig. 7(B2)) is much larger than that of the adjustment applied to the first stroke (Fig. 7(B3)). This confirmed the experts' knowledge. The experts explained that this is because the influence of the first stroke on the final result is limited. Thus, no matter which attributes we adjust at the first stroke, the final winning rate will not be obviously affected. Hence, the experts paid attention to the other optimum strategy (Fig. 7(B2)). They found that the original scoring rate of the optimum adjustment at the third stroke (Fig. 7(B1)) is not the highest among all adjustments with the same technique (Fig. 7(B4)).

This is a new insight for the experts. Therefore, they added the adjustment with the highest scoring rate to the evaluation component (Fig. 7(B5)). They noticed that the effect (winning rate) of the newly added adjustment (Fig. 7(B2) is almost the same as the optimum one (Fig. 7(B5)). However, considering the adjustability, they eventually chose the optimum adjustment.

To figure out the influence and the implementation of the optimum adjustment, the experts turned to the explanation mode. They noticed that this adjustment mostly affects the fifth stroke (the solid line to the fifth stroke is thicker than that of the fourth line (Fig. 7(C2)). In the fifth stroke, Ito is most likely to use quick attack to match the quick attack of the adjustment (Fig. 7(C5)). The experts further clicked the quick attack in the fifth stroke and found the placement of the strokes hit by quick attack is also long backhand, the same as that of the adjustment (Fig. 7(C3)). They commented that this conclusion could hardly be obtained in the past without this system. They further concluded that this adjustment improves Ito's performance by increasing Ito's tactic, **consecutive quick attack to long backhand**. This tactic should be one of Ito's dominant tactics. As for the implementation, the experts inspected the second stroke (the dashed line from the second stroke is thicker than that from the first stroke (Fig. 7(C1)). As Fig. 7(C4) shows, the adjustment can be achieved mostly after Ito's opponents employ offensive techniques (i.e., topspin and flick) at the second stroke. According to the experts, this suggests that Ito can still take back the initiative through the dominant tactic even her opponents attack at first, which matched the experts' knowledge about Ito' s playing style.

Through this case study, the experts concluded that coaches can focus on training Ito's tactic, **consecutive quick attack to long backhand** at the third stroke of her serve rallies to improve her performance.

6.3 Insight 2: Topspin to Long Forehand is the Key to Receive Rallies for Ding Ning.

This case study mainly focuses on improving Ding Ning's performance by adjusting her receive rallies. Ding Ning is the top table tennis player in the world and thus, the experts are also interested in her matches.

The experts chose Ding Ning to examine her matches. Since there are only 7 matches of Ding Ning, they decided to analyze all matches of Ding Ning. In the tactic view, they also first examined Ding's tactics in serve rallies from the perspective of the stroke technique. Similarly, they sorted the tactics based on the scoring rates and only showed the tactics whose utilization rates are more than 1%. As Fig. 8(A) shows, the scoring rates of all tactics are proportional to their utilization rates and most of the scoring rates are larger than 50%. The experts agreed with this result and explained that Ding performs well during her serve phase, thus, there is little space to improve Ding's performance in her serve phase. Thereafter, they turned to Ding's receive phase of receive rallies with the same sorting and filtering conditions. As Fig. 8(B) shows, Ding's tactics in the receive phase are worse compared to those in the serve phase since the scoring rates of most of her tactics here are less than 50% (Fig. 8(B2)). This result also matched the experts' knowledge about Ding Ning. They explained that Ding often uses push, a control technique while receiving the serve of her opponents, which means Ding often fails to attack at first and loses the initiative. They decided to take these tactics as a breakthrough to improve Ding's performance and chose all these tactics.

The experts activated the stroke placement in the simulation view and let the system generate the optimum strategies at first. According to the strategies consisting of two adjustments (Fig. 8(C)), if Ding can use topspin, an offensive technique, while receiving the serve and keep using topspin afterward more often, she can significantly enhance her winning rates. The experts agreed with this strategy since consecutive topspin within receive rallies is the tactic with highest scoring rates (Fig. 8(B1)). However, the optimum strategy consisting of a single adjustment at the second stroke (Fig. 8(D)) confused them. There is an adjustment that shares the same stroke technique and scoring rate with the optimum one and has higher adjustability in the option list (Fig. 8(E, E1)) but it is not the optimum adjustment. To figure out the reason, they added this adjustment to the evaluation component.

In the explanation component, the experts examined the influences of the two adjustments separately. They noticed that the differences between the two adjustments mainly lie in their individual influences to

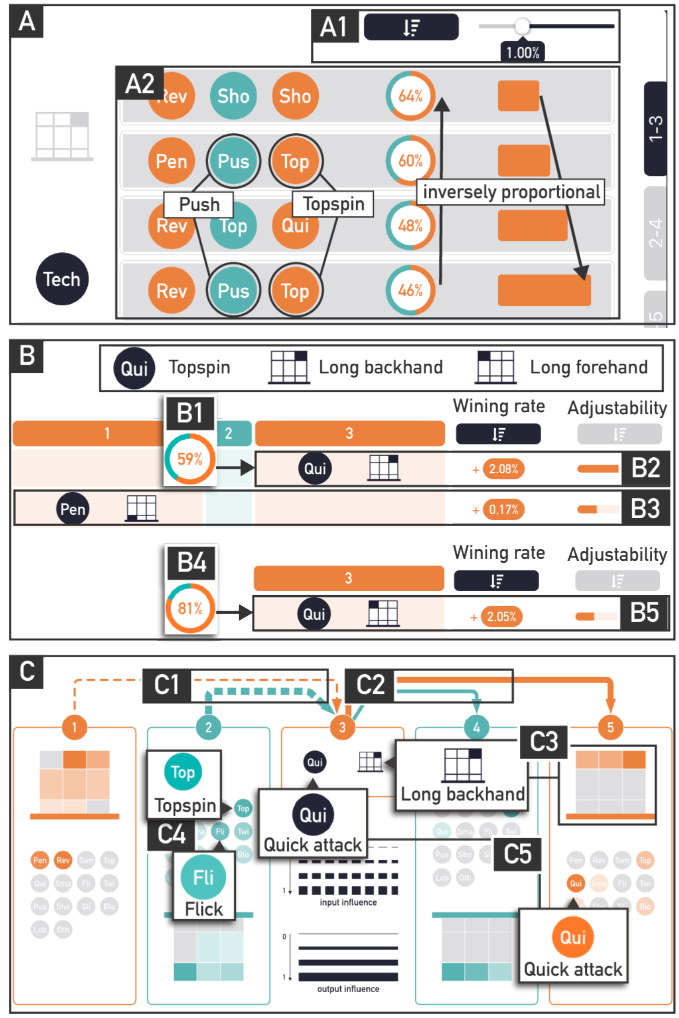

Fig. 7. The figure of Section 6.2. (A) presents the tactics used by Ito Mima in the serve phase. (B) presents the adjustment strategies in the evaluation component. (C) presents explanation of the adjustments.

the third stroke. If Ding hits the ball to long middle, then her opponents will hit back to long middle ((Fig. 8(E2)) while if Ding hits the ball to long forehand, then her opponents will hit back to long forehand (Fig. 8(D2)). This is a new insight for the experts. The experts explained that the power of topspin of Ding can be enhanced if it is used to receive the ball hit to forehand. Therefore, the stroke hit to long forehand will perform better than that hit to long middle.

Through this case, the experts concluded that coaches can focus on training Ding to receive her opponents' serve by **topspin to long forehand** to improve her performance.

6.4 Expert Feedback

We interviewed the experts individually after each case study and summarized their feedback as follows.

Usability. According to the experts, Tac-Simur significantly facilitates their analysis. Expert A appreciated the exploration mode most because the combination of multiple stroke attributes within the adjustments expands their exploration space. As for expert B, the visual explanation of the simulation process impresses him most. He could figure out not only why particular adjustment performs well/poorly, but also how to achieve a particular adjustment. However, the experts also mentioned two limitations of Tac-Simur during the case studies. First, the icon of stroke position is not friendly for all users. One of the experts often forgot the exact meaning of each icon due to his habit of swinging the racket which is different from that of others. He often asked our experimenter for confirmation. After using the system for a while, He remembered the encodings eventually. The second is in-

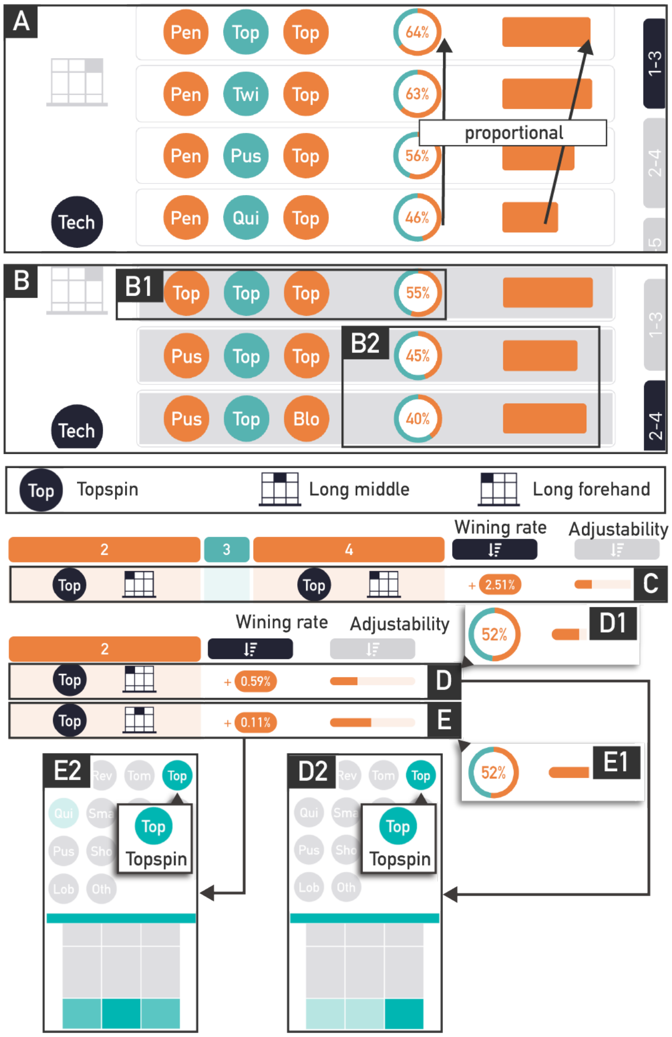

Fig. 8. The figure of Section 6.3. (A) presents the tactics used by Ding Ning in the serve phase and (B) presents those in the receive phase. (C), (D), and (E) present adjustment strategies in evaluation component. (F) and (G) present the explanation of the adjustments.

teractions in the explanation mode. The affordance of interactions for correlating stroke attributes is poor. The experts often investigated the correlation among different attributes without correlating them.

Suggestion. The experts proposed two suggestions about Tac-Simur. First, the experts hoped to add more stroke attributes to the match simulation. Second, for more detailed analysis, the experts hoped to access the raw videos of the matches to further verify their analysis.

7 Discussion

In this section, we discuss the significance, generalizability, lessons learned and limitations of our work.

Significance. Simulative analysis plays an important role in many fields [42, 43], especially in competition sports (e.g., basketball [31] and tennis [35]). We introduce a visual analytics system, Tac-Simur, to facilitate simulative analysis of table tennis. With Tac-Simur, the experts have discovered valuable strategies to improve players' performance in future matches. These strategies are appreciated and accepted by the coaches in one of the top national table tennis teams in the world.

Generalizability. Tac-Simur can be extended to other sports that can be simulated by our model. Our model is intrinsically a state-transition model, therefore, it can be extended to other similar sports whose processes can also be quantified as transitions between different states. For example, in tennis, each shot can be treated as the state, and the interaction between the shots of the players can be treated as the transition process. However, as for sports like soccer and basketball, our model can hardly be extended to them. The rules are more complicated and there are more players, which means more factors should be considered for simulation of such sports. The definition of states and simulation process of our model is too simple for such complicated sports. We hope to improve the generalizability in the future.

Lessons learned. We have learned three lessons through this design study. First, for the problem domain, a pilot system is necessary. Abstract domain knowledge and requirements are difficult to understand only through meetings. A simple pilot system developed according to existing methods can also work as a Liaison [29] to bridge the gap between us and the experts. It can facilitate the process of identifying the problem domain with the experts. Second, for the simulative analysis, the adjustments to the variables should be restricted. Initially, we allowed the experts to change the utilization rates of the strokes to any values. However, such flexible exploration made the experts spend abundant time on the trade-off between the effect of the adjustment and the possibility of achieving the adjustment in practice. Therefore, we provide restricted amounts of the changes to utilization rates of the strokes which are pre-computed based on Equation. 2 during exploration. Third, for the visualization, simple visual design has higher legibility and generalizability. The major visual elements in Tac-Simur are widely-used donut charts and bar charts instead of complicated tailored visual encodings. Therefore, the experts can easily understand the interface. Furthermore, the visualization of Tac-Simur can be easily extended to other similar sports.

Limitations. The limitations of our work mainly lie in three aspects. First, our dataset is too small to conduct analysis on more players. The training of the new model usually requires data of more than five matches since more stroke attributes are included. However, for most of the players in our dataset, we only have the data of one or two matches. Therefore, in the future, we will collect more data of different players. Second, the model only considers the three most significant attributes, namely, stroke position, stroke techniques, and stroke placement while ignoring several other stoke attributes, such as stoke effect, stoke spin, and stoke action. There is still much room for improving the model to make it more comprehensive by integrating more attributes. However, such integration may lead to sparse matrices and vectors. In the future, we plan to explore how to integrate more attributes into the model while solving the spareness issue. Third, the visual design of the explanation mode cannot display multiple adjustments together. One adjustment is related to four strokes, thus the relationships among an entire adjustment strategy are complicated and difficult to visualize clearly. However, only presenting one adjustment of a strategy limits the comprehension of the whole strategy. In the future, we plan to explore and study how to visualize a whole strategy in one view clearly.

8 Conclusion

In this work, we identify the problem domain of tactic-based simulative visual analytics of table tennis. We subsequently introduce a hybrid second-order Markov chain model to simulate matches more adequately. Further, based on the model, we develop a visual analytics system named Tac-Simur to help facilitate simulative analysis.

In the future, we plan to improve this work from two aspects. First, we will try to expand our dataset. On the one hand, we will continue collecting data from various events such as the upcoming *Liebherr 2019 ITTF World Table Tennis Championships*. On the other hand, we plan to deploy our systems including a data collection system online for table tennis teams to use. In this manner, we can include not only the match data but also the training data. Second, we will try extending our model to other sports like tennis, badminton, and even soccer. We hope to construct a more general model that can simulate most of the ball games in competitive sports. We hope to figure out the features shared by ball games during simulative analysis and provide general guidelines for simulative visual analytics of ball games.

Acknowledgement

The work was supported by National Key R&D Program of China (2018YFB1004300), NSFC (61761136020), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1609217), Zhejiang Provincial Natural Science Foundation (LR18F020001) and the 100 Talents Program of Zhejiang University.

References

[1] R. Y. S. Aoki, R. M. Assunção, and P. O. S. V. de Melo. Luck is Hard to Beat: The Difficulty of Sports Prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1367–1376, 2017.

[2] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry. POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. In *Proceedings of MIT Sloan Sports Analytics Conference*, 2014.

[3] S. Chen and T. Joachims. Predicting Matchups and Preferences in Context. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 775–784, 2016.

[4] W. Chen, T. Lao, J. Xia, X. Huang, B. Zhu, W. Hu, and H. Guan. GameFlow: Narrative Visualization of NBA Basketball Games. *IEEE Transactions on Multimedia*, 18(11):2247–2256, 2016.

[5] W. K. Ching, E. S.Fung, and M. K. Ng. Higher-Order Markov Chain Models for Categorical Data Sequences. *Naval Research Logistics*, 51(4):557–574, 2004.

[6] P. O. S. V. de Melo, V. A. F. Almeida, A. A. F. Loureiro, and C. Faloutsos. Forecasting in the NBA and Other Team Sports: Network Effects in Action. *ACM Transactions on Knowledge Discovery from Data*, 6(3):13:1–13:27, 2012.

[7] C. A. Dietrich, D. Koop, H. T. Vo, and C. T. Silva. Baseball4D: A Tool for Baseball Game Reconstruction and Visualization. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 23–32, 2014.

[8] M. Fuchs, R. Liu, I. M. Lanzoni, G. Munivrana, G. Straub, S. Tamaki, K. Yoshida, H. Zhang, and M. Lames. Table Tennis Match Analysis: A Review. *Journal of Sports Sciences*, 36(23):2653–2662, 2018.

[9] A. Gabel and S. Redner. Random Walk Picture of Basketball Scoring. *Journal of Quantitative Analysis in Sports*, 8(1):1–20, 2012.

[10] K. Goldsberry. Courtvision: New Visual and Spatial Analytics for the NBA. In *Proceedings of MIT Sloan Sports Analytics Conference*, 2012.

[11] M. Lage, J. P. Ono, D. Cervone, J. Chiang, C. A. Dietrich, and C. T. Silva. StatCast Dashboard: Exploration of Spatiotemporal Baseball Data. *IEEE Computer Graphics and Applications*, 36(5):28–37, 2016.

[12] M. Lames. *Leistungsdiagnostik durch Computersimulation: ein Beitrag zur Theorie der Sportspiele am Beispiel Tennis*. Deutsch Harri GmbH, Frankfurt, 1991.

[13] I. M. Lanzoni, R. D. Michele, and F. Merni. Technical and Tactical Performance of Top-Class Senior, Junior and Cadet Table Tennis Players. *International Journal of Table Tennis Sciences*, 8:80–85, 2013.

[14] P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis. *Computer Graphics Forum*, 31(3):1255–1264, 2012.

[15] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):1–10, 2017.

[16] A. G. Losada, R. Therón, and A. Benito. BKViz: A Basketball Visual Analysis Tool. *IEEE Computer Graphics and Applications*, 36(6):58–68, 2016.

[17] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski. The State-of-the-Art in Predictive Visual Analytics. *Computer Graphics Forum*, 36(3):539–562, 2017.

[18] S. Merritt and A. Clauset. Scoring Dynamics Across Professional Team Sports: Tempo, Balance and Predictability. *EPJ Data Science*, 3(1):4, 2014.

[19] A. W. Y. Ng and A. H. S. Chan. Visual and cognitive features on icon effectiveness. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 2, pages 19–21, 2008.

[20] J. P. Ono, C. A. Dietrich, and C. T. Silva. Baseball Timeline: Summarizing Baseball Plays Into a Static Visualization. *Computer Graphics Forum*, 37(3):491–501, 2018.

[21] L. Peel and A. Clauset. Predicting Sports Scoring Dynamics with Restoration and Anti-Persistence. In *Proceedings of IEEE International Conference on Data Mining*, pages 339–348, 2015.

[22] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013.

[23] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018.

[24] M. Pfeiffer, H. Zhang, and A. Hohmann. A Markov Chain Model of Elite Table Tennis Competition. *International Journal of Sports Science and Coaching*, 5(2):205–222, 2010.

[25] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko. SnapShot: Visualization to Propel Ice Hockey Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2819–2828, 2012.

[26] T. Polk, J. Yang, Y. Hu, and Y. Zhao. TenniVis: Visualization for Tennis Match Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2339–2348, 2014.

[27] F. Pradas, P. Martínez, M. Rapún, V. Bataller, C. Castellar, and L. Carrasco. Assessment of Table Tennis Temporary Structure. In *Proceedings of the 12th ITTF Sports Science Congress*, pages 80–85, 2011.

[28] A. E. Raftery. A Model for High-Order Markov Chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):528–539, 1985.

[29] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. Bridging the gap of domain and visualization experts with a liaison. In *Accepted at the Eurographics Conference on Visualization (EuroVis 2015, Short Paper)*, volume 2015, pages 127–131. The Eurographics Association, 2015.

[30] S. Tamaki, K. Yoshida, and K. Yamada. A Shot Number Based Approach to Performance Analysis in Table Tennis. *Journal of Human Kinetics*, 55(1):7–18, 2017.

[31] P. Vračar, E. Štrumbelj, and I. Kononenko. Modeling basketball play-by-play data. *Expert Systems with Applications*, 44:58–66, 2016.

[32] W. Wang, J. Zhang, X. Yuan, and S. Liu. Matchorchestra: a generalized visual analytics for competitive team sports. *Journal of Visualization*, 19(3):515–528, 2016.

[33] X. Wei, P. Lucey, S. Morgan, P. Carr, M. Reid, and S. Sridharan. Predicting Serves in Tennis Using Style Priors. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2207–2215, 2015.

[34] X. Wei, P. Lucey, S. Morgan, and S. Sridharan. Predicting Shot Locations in Tennis Using Spatiotemporal Data. In *Proceedings of 2013 International Conference on Digital Image Computing: Techniques and Applications*, pages 1–8, 2013.

[35] X. Wei, P. Lucey, S. Morgan, and S. Sridharan. "Sweet-Spot": Using Spatiotemporal Data to Discover and Predict Shots in Tennis. In *Proceedings of MIT Sloan Sports Analytics Conference*, 2013.

[36] D. Weng, R. Chen, Z. Deng, F. Wu, J. Chen, and Y. Wu. Srvis: Towards better spatial integration in ranking visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):459–469, 2018.

[37] D. Weng, H. Zhu, J. Bao, Y. Zheng, and Y. Wu. Homefinder revisited: Finding ideal homes with reachability-centric multi-criteria decision making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 247:1–247:12, 2018.

[38] S. Wenninger and M. Lames. Performance Analysis in Table Tennis - Stochastic Simulation by Numerical Derivation. *International Journal of Computer Science in Sport*, 15(1):22–36, 2016.

[39] H. Wu, Z. Li, Z. Tao, W. Ding, and J. Zhou. Methods of Actual Strength Evaluation and Technical Diagnosis in Table Tennis Match. *Journal of National Research Institute of Sports Science*, 1:32–41, 1989.

[40] Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, and H. Zhang. iTTVis: Interactive Visualization of Table Tennis Data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):709–718, 2018.

[41] Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, S. Cheng, and W. Chen. ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):65–75, 2019.

[42] M. Xu, Z. Pan, M. Zhang, P. Lv, P. Zhu, Y. Ye, and W. Song. Character behavior planning and visual simulation in virtual 3d space. *IEEE Multimedia*, 20(1):49–59, 2013.

[43] M. Xu, H. Wang, S. Chu, Y. Gan, X. Jiang, Y. Li, and B. Zhou. Traffic simulation and visual verification in smog. *ACM Transactions on Intelligent Systems and Technology*, 10(1):3:1–3:17, 2019.

[44] H. Zhang. *Leistungsdiagnostik im Tischtennis*. Verlag Dr. Kovac, Hamburg, 1 edition, 2006.

[45] H. Zhang, W. Liu, J. Hu, and R. Liu. Evaluation of Elite Table Tennis Players' Technique Effectiveness. *Journal of Sports Sciences*, 31(14):1526–1534, 2013.

[46] H. Zhang, L. Yu, and J. Hu. Computer-aided Game Analysis of Net Sports in Preparation of Chinese Teams for Beijing Olympics. *International*

Journal of Computer Science in Sport, 9(3):53–69, 2010.
[47] H. Zhang and Z. Zhou. Two-factor combination analysis model of the first four strokes for elite shake-hand table tennis players and its application. *Journal of Shenyang Sport University*, 36(6):115–124, 2017.
[48] M. Zhang, L. Chen, X. Yuan, R. Huang, S. Liu, and J. Yong. Visualization of technical and tactical characteristics in fencing. *Journal of Visualization*, 22(1):109–124, 2019.